Selection as a Factor in Human Survivorship over the Past Three Centuries

Hansen, Hans Oluf

Publication date:
2011

Document version
Early version, also known as pre-print

Citation for published version (APA):
Hansen, H. O. (2011). *Selection as a Factor in Human Survivorship over the Past Three Centuries*. Paris: International Union for the Scientific Study of Population.

Prepared for the Seminar on Lifespan Extension and the Biology of Changing Cause-of-Death Profiles: Evolutionary and Epidemiological Perspectives, organized by the IUSSP Scientific Panel on Evolutionary Perspectives in Demography, Rauischholzhausen, Germany, 13-15 January 2011.

Selection as a factor in human survivorship
Over the past three centuries

Hans O. Hansen

Assoc. Prof. (Demography) (Emeritus)

University of Copenhagen, Dept. of Economics

5 Øster Farimagsgade 5, Building 26.0.24

DK-1353 Copenhagen K, Denmark

Phone +45 35 32 32 61

E-mail HansOluf.Hansen@econ.ku.dk

Home page http://www.econ.ku.dk/personal/usihoh/

Ftp-server: ftp://ftp.ibt.ku.dk/usihoh/

Date of last revision: 27 March 2011

Abstract

This study reviews an explanatory statistical model targeting cause-aggregated cohort mortality by *sex* and *age* as a function of individual genetics; environmental factors with/without impact for selection; and a latent age-specific non-parametric baseline hazard joint for all individuals. The baseline, indicating degenerative biological ageing, is identified up to a multiplicative factor. A summary of results on fitting the model to historical cohort data covering extreme variation in empirical mortality is considered. The explanatory model does not claim to be fully accurate; however, its relevance is justified by its documented capacity to fit a substantial portion of the knowledge, observations and theoretical circumstances about human survivorship over the past two or three centuries, in particular prior to the latter decades of the 19th century when improved prophylaxis and artificial immunization became increasingly widespread with the progress of sanitation and advances of medical technology. Extension of the model to address morbidity by diagnosis and mortality by age and cause associated with selection and environmental impact is discussed and illustrated by an example.

1. Introduction

Historical mortality change has profound impact on current cross-sectional mortality because of selection rooted in individual genetics and personal survivor experience (Hansen 2008). In the modern world hardly any population is unaffected of selection associated with the demographic transition. Long-term mortality change is on reasonably consistent and reliable historical record in a few societies.

2. A simple life model with selection and environmental impact

To sort out structural from stochastic elements in historical mortality change we search for a model accommodating and respecting a few empirically obvious conditions. First, people are neither genetically equal, nor do they experience identical risks or exposure to life course events. Second, life is finite, whatever its upper limit. Third, differential mortality at the levels of individuals has selection effects; this is the famous survival-of-the-fittest principle coined by Herbert Spencer (1820-1903) but usually attributed to Charles Darwin (1809-1882). To minimize heterogeneity, inference on survivorship must be based on individual life courses or aggregate data related to birth cohorts.

Modern highlights in the history of demographic and statistical modeling of heterogeneity in human survivorship include contributions of Gompertz (1825), Makeham (1860), Cox (1972), and Vaupel et al. (1979). The Makeham law extends the Gompertz function by introducing an additive death risk which is independent of age. The Gompertz part of the Gompertz-Makeham law describes mortality as an exponential function of age. The Cox model broadens this view by introducing a shared unspecified baseline intensity depending on time or age; multiplicatively related to an exponential function admitting multivariate personal characteristics which may be fixed or vary over time.

An extension of multiplicative hazard modeling of the role of genetic heritage is due to Vaupel et al. (1979), who depict mortality as a multiplicative function of congenital biological frailty at the level of individuals and a joint baseline hazard depending on age. Several important questions are left open. For example, how do environmental factors during gestation influence the frailty distribution of live births in a cohort? Moreover, Vaupel et al. (1979) neither consider choice or interpretation of their baseline mortality, nor do they reflect on environmental interaction with human survivorship

after conception and live birth. In weighty empirical work Bourgeois-Pichat (1946, 1951) singles out so-called *endogenous* and *exogenous* elements in infant mortality. The embedded statistical model is basically competing risks. An essentially descriptive statistical approach leads Bourgeois-Pichat (1951) to important conclusions. First, endogenous biological factors dominate neonatal mortality while post-neonatal mortality is governed by exogenous environmental factors. Second, in populations with general access to steadily improved medical technology the endogenous element has emerged as the principal determinant of infant mortality. In old first-world countries this development dates back to the great medical advances by distinguished researchers such as Pasteur, Koch, and Lister in the late nineteenth century followed by Fleming and Salk in the early and mid-twentieth century. Like many earlier statisticians and demographers, Bourgeois-Pichat ignored natural selection. Commonplace in actuarial reckoning the notion of selection only occasionally has found its way into population studies and demography. For a Danish example cf. Westergaard (1898).

2.1 Model

The basic life model is defined on state space S_0 = {Alive, dead}. With *x* denoting age and z_υ indicating individual frailty the survival of individual υ is governed by death risk $m_\upsilon(z_\upsilon, x)$.

Consider the following hazard model.

$$m_\upsilon(z_\upsilon, x) = z_\upsilon(x=0)\, m_s(z=1, x)\, \varepsilon_t + \theta_t \qquad (1.1)$$

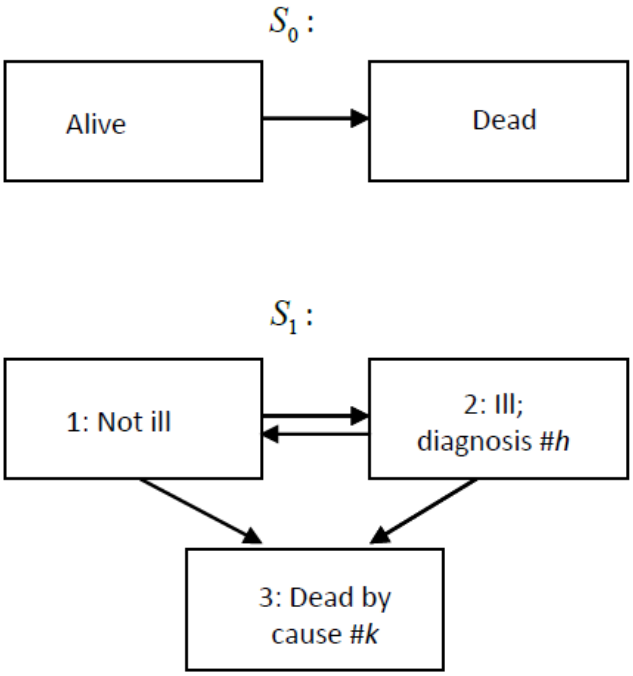
Where

$$Z_\upsilon \sim Gamma(\alpha, \beta, \gamma = 0)$$

Model (1.1) embodies a competing risks mechanism: at time t individual υ may die either from risk $\left(z_\upsilon(x=0)\, m_s(z=1, x)\, \varepsilon_t\right)$ or from risk θ_t; the latter referring to sudden death, for example caused by natural disasters such as a tsunami or an earthquake and having no selection effects. Under the competing risks model the two types of death risk are stochastically independent. The gamma variate Z_υ denotes individual frailty of person υ. Statistics ε_t and θ_t indicate exogenous impact on mortality respectively with/

without selection among survivors at time t; and $m_s(z=1,x)$ represents a baseline hazard shared by survivors aged x at time t. As the survivors differ by personal frailty, so does their mortality: individuals with high frailty have higher mortality than individuals with low frailty. The frailty distribution $Z(x)$ of cohort survivors vary with age x due to selection as the cohort get trimmed of individuals unfit, for one reason or another, to staying alive.

Figure 1 State spaces S_0 and S_1



2.2 Estimation

Evaluating the capability of model (1) to describe empirical cohort mortality on reliable historical record entails fitting the model to schedules ranging from very high unbounded to very low highly controlled cohort mortality, including schemes of transition from high-to-low mortality over chosen "transitory" birth cohorts, in the first place for populations and epochs where environmental mortality θ_t without impact for selection may safely be set to nil. For an outline of the empirical challenge to model-based mortality research cf. Figure A1, Appendix. Once the baseline hazard

$m_s(z=1,x)$ has been identified the model may be applied to evaluating θ_t in populations where this is relevant.

Ideally age x represents time elapsed since conception. However, data limitations rarely allow evaluating the frailty distribution prior to live birth. The live births of a cohort will already have been depleted by miscarriages and natural abortion during gestation. The variance of the frailty distribution on live birth will be smaller than, and therefore not representative of the variance of the frailty distribution $z(x)$ on conception. This is because of selection during gestation. Under model (1.1) the survivors $\ell(x,z(x))$ are purged randomly over the entire life span $0<x<\omega$, $\omega=\max(x)$, of individuals with relatively high frailties.

Fitting model (1.1) to empirical cohort mortality naturally entails minimizing the deviation between empirical cohort mortality $m(x)$ and predicted (model-based) mortality $\tilde{m}(z,x)$ with regard to gamma parameters α, β and baseline $m_s(z=1,x)$. For unique identification two of the factors $z(x)$, $m_s(z=1,x)$, ε_t need to be normalized. Heterogeneity of the risk population clearly rules out maximum likelihood estimation of baseline hazard $m_s(z=1,x)$. These conditions suggest the following strategy.

First assess ε_t by fitting a log-linear intensity model to cross-sectional occurrence/exposure data by year, age, and possibly sex covering the reference period $[t_0,t_1[$. Next, normalize ε_t by dividing with ε_{t_r} i.e. $\varepsilon_t=\varepsilon_t/\varepsilon_{t_r}$, $t_r\in[t_0,t_1[$, t_r denoting birth year of the cohort; this makes $\varepsilon_{t_r}=1$. The mean frailty may then be evaluated as $z(x=0)=m_{t_r}(x=0)/m_s(z=1,x=0)$, age x here denoting time elapsed since live birth. With the data available we use empirical infant mortality to obtain a preliminary assessment of statistic $m_{t_r}(x=0)$. Heterogeneity is basically determined by the variance of the distribution of congenital frailty. What variance of the frailty distribution on live birth should be deployed to minimize the squared deviation between empirical and predicted aggregate mortality while identifying the baseline hazard and respecting empirical structural traits? Can a joint baseline hazard be found? Is degenerative

biological ageing as indicated by the baseline hazard, independent of *sex* and *birth cohort*?

To answer these questions a very large number of computerized trials was required. For some highlights of results cf. Section 2.3. Estimation of model-based cohort mortality $\tilde{m}(z,x)$ rests on stochastic micro-simulation in the simple life model with the state space S_0 ={Alive, dead}. See Hansen (2000, 2008) for further details.

2.3 Some highlights of results on fitting model (1.1) to empirical cohort data

Hansen (2008) actually identifies unique solutions, not only of the gamma parameters but also of the otherwise unobservable non-parametric baseline $\tilde{m}_s(z=1,x), 0 \le x < \omega$.

The description of mortality by model (1.1) turns out to be rather general and of a quality second to none based on parametric statistical approaches, for example Gompertz (1825), Makeham (1860), Vaupel et al. (1979), Lee and Carter (1993), and surely many others. With the above-mentioned normalization the identified baseline $\tilde{m}_s(z=1,x)$, interestingly, turns out to be joint for men and women across the empirical birth cohorts considered. Due to unknown selection during gestation the identified sets of gamma parameters on live birth exhibit considerable dependency on the timing of birth of the cohorts; with obvious consequences for professional understanding of biological and ecological elements in survivorship. A statistical summary of results is shown in Tables A1 and A2, Appendix. Detailed documentation of graphical control on fitting model (1.1) to empirical mortality of the elected birth cohorts may be found at ftp://ftp.ibt.ku.dk/usihoh/Selection in human survivorship/.

In the following we illustrate, briefly, the capacity of our model to approach empirical cohort mortality subject to extreme variation. We look into the recovered baseline hazard representing biological ageing along with shape and form of the recovered probability distributions representing latent congenital frailty. The model offers ample support to the explanation by Westergaard (1898) of selection as a natural and coherent general explanation of the somehow paradoxical stagnation or transitory deceleration of cohort mortality in the extreme ages, from around age 92 and beyond, say. We illustrate this phenomenon by a couple of examples. By the model-based analysis it seems fair to conclude that selection has tremendous impact on human health indicated by the

Figure 2

Swedish females born 1751: Empirical and model-based mortality (per 10,000)

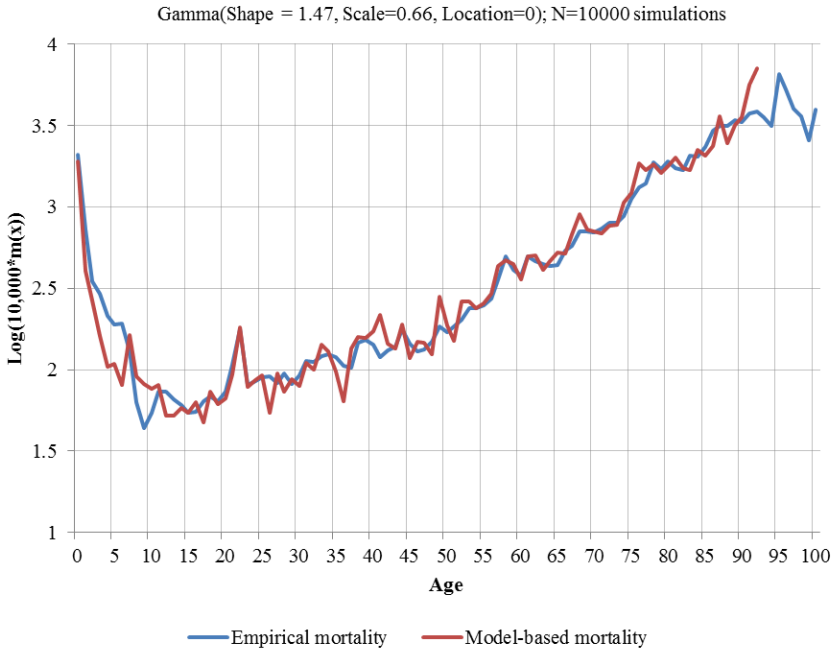


Figure 3

Swedish females born 1801: Empirical and model-based mortality (per 10,000)

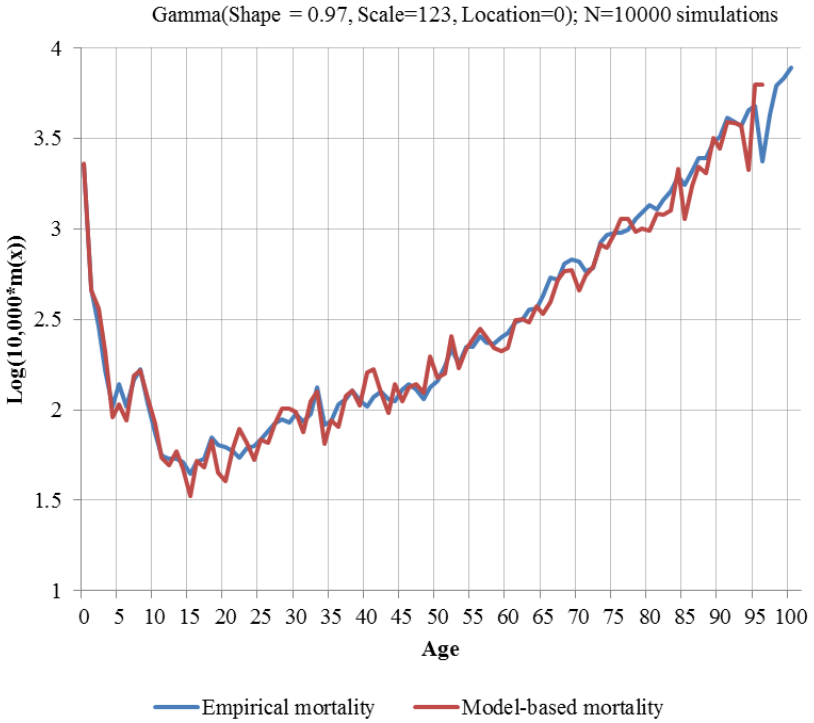


Figure 4
Danish females born 1901: Empirical and model-based mortality (per 10,000)

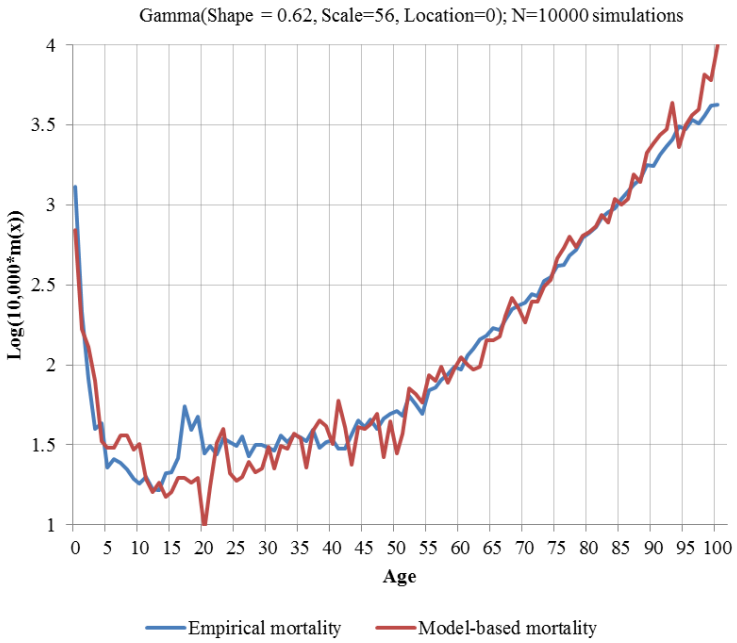
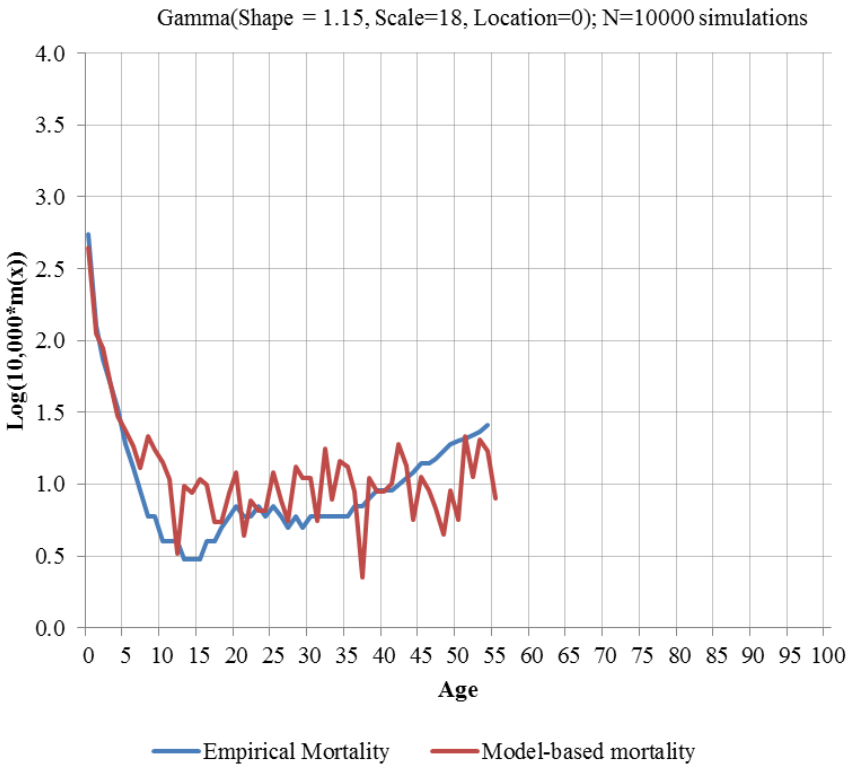


Figure 5 Japanese females born 1950: Empirical and model-based mortality (Per 10,000)



changing frailty composition of survivors in the course of the modern demographic transition. We illustrate selection spurred by multifactor infections and misery during the environmental disturbances 1783-84 and the human catastrophe in Iceland 1784-86. We also consider some demographic effects of the Spanish Flue around 1918.

2.3.1 Some highlights of the graphic model control

To illustrate significant long term mortality change in three current First-World countries figures 2-5 display empirical and fitted cohort mortality among elected female cohorts born in Sweden 1751 and 1801, in Denmark 1901, and in Japan 1950 i.e. cohorts born before, during, and after the modern long term mortality decline cf. Figure A.1, Appendix. The examples show that model (1.1) in general approaches empirical mortality extremely well, both in traditional and modern societies. The model description of the Danish female cohort born in 1901 does not capture the Spanish Flue too well. This is because of a failing multiplicative relationship between *age* and *time* in mortality during World War I. For unknown reasons, perhaps acquired immunity from past influenza epidemics, crisis mortality was markedly higher in the ages below 35 than in mature and elderly ages; leading to undervaluation of trend $\varepsilon(t), t \in [1914,1918]$.

2.3.2 Variation in latent congenital frailty in the past two to three centuries

There are two latent elements in the model viz. the baseline hazard representing biological ageing as a function of age; and congenital frailty on live birth. Figure 6 exhibits the recovered biological baseline which is common for men and women. Although the model offers quite satisfactory fits to empirical mortality (figures 2-5) in all ages there may be some scope for improvement of the baseline hazard, not least among infants and young children. This would likely involve working with age intervals smaller than one year in the baseline hazard. A strengthening of this part would also involve additional focus on gestational survivorship in the manner of Bourgeois-Pichat (1951, 1952) and Hansen (1982a-b, 1989, 1996). The empirical mortality experience of the populations considered in this study does not permit secure recovery of the latent biological baseline beyond age 90.

Figure 6

Recovered and guessed latent biological baseline hazard

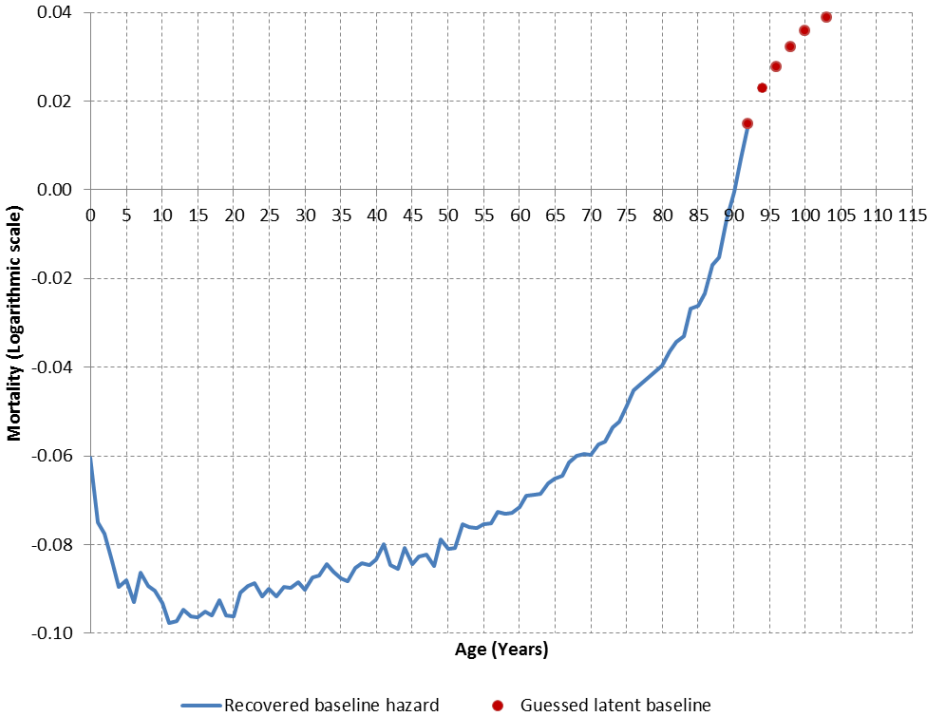
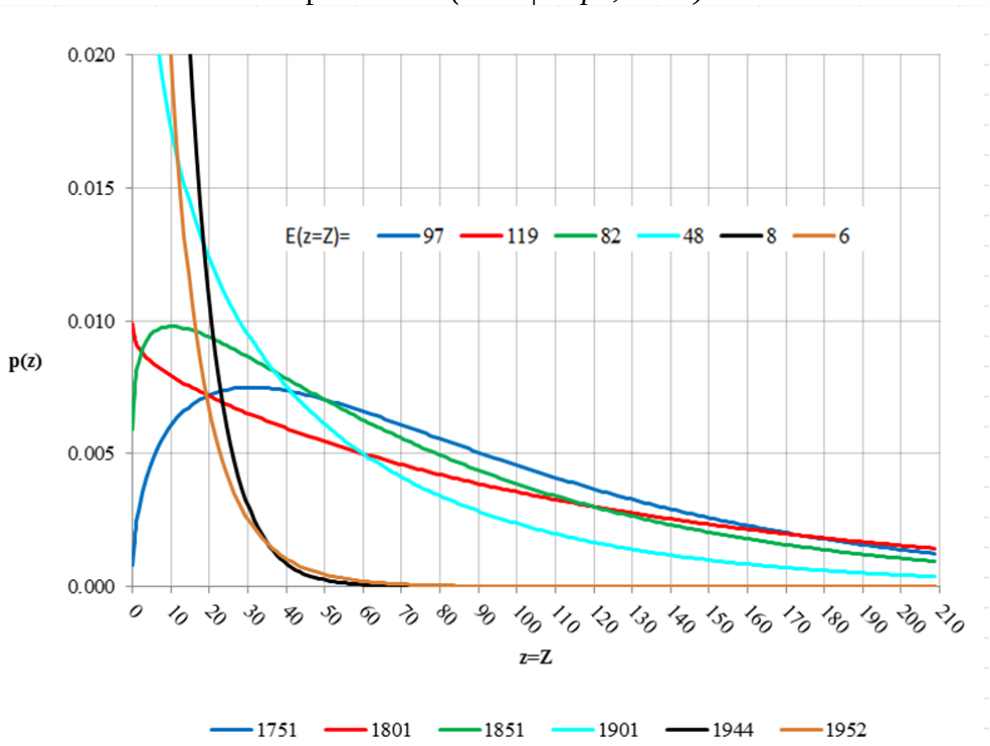


Figure 7

Gamma probability densities *p(z|shape,scale)* of Swedish females by birth cohort and expectation *E(z = Z|shape, scale)*



Note. Cf. Table A.2 for detailed listing of the actual shape and scale values of the distributions

That health or susceptibility to illness differs across people and age is plain to anyone. Personal frailty or susceptibility to illness ultimately leading to death may be difficult to diagnose, not least during gestation and on birth of live children. The notion of latent *congenital frailty* seems meaningful, hence. Figure 7 graphs, as an example, the gamma probability densities $p(z \mid shape, scale)$ against personal frailty z on fitting the model to empirical mortality of the elected female Swedish birth cohorts; with an outline of the corresponding mean frailties. Figure 7 reiterates general results exhibited in Table1 A.1-2. The mean and variance of latent congenital frailty required to obtain close model-based fits to extreme variation in empirical mortality have diminished dramatically over time as the gap between empirical cohort mortality and the latent baseline has become smaller. Being an indicator of the general level of reproductive health this change is intimately related to the development of medical technology and know-how in the course of the demographic transition. That reproduction might instigate selection across birth cohorts to influence frailty on conception could well be an issue in evolutionary history but hardly over two to three centuries considering life length and distances between generations in the human species. Further discussion of this question is outside the scope of the present study.

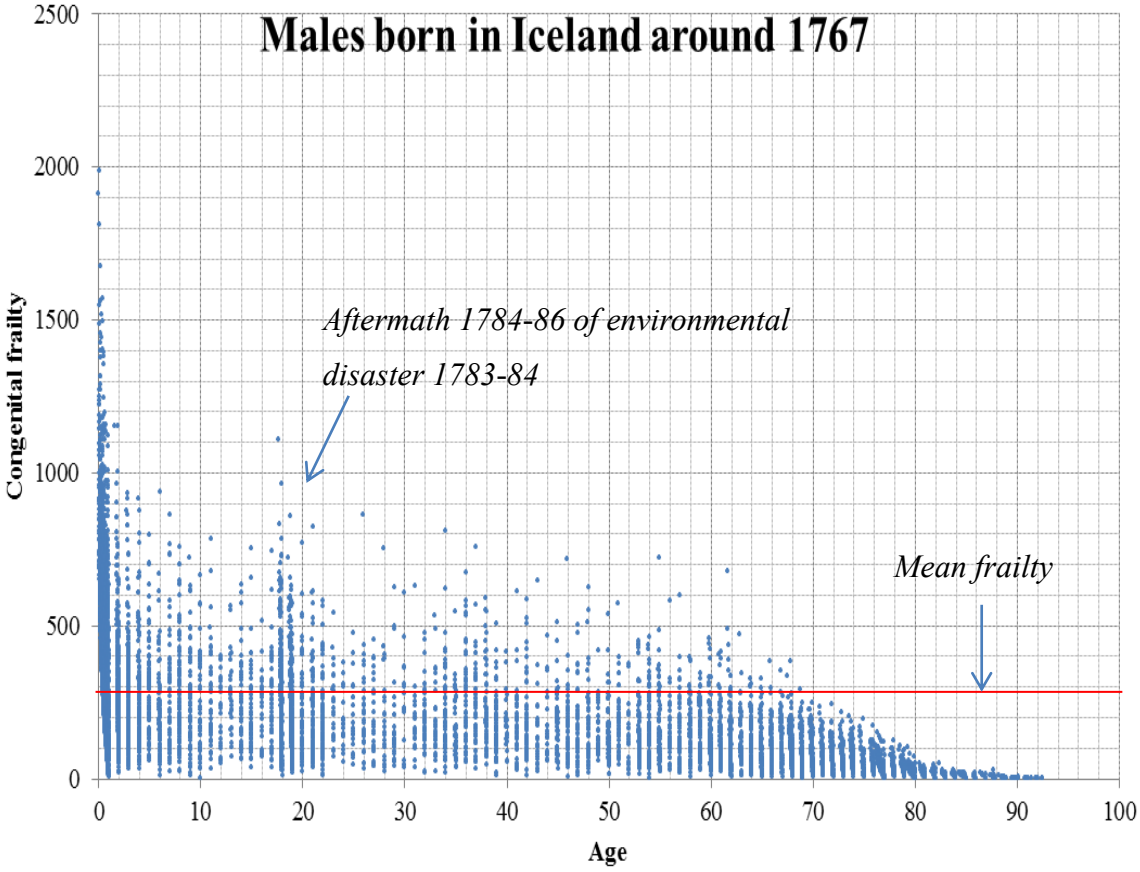
From a statistical perspective there is nothing salient about the choice of probability distribution of the congenital frailties on livebirth. Drawing individual frailties from some initial probability distribution and multiplying it to a common age-dependent baseline hazard is just a straightforward way of personalizing survivorship. Other probability distributions than the gamma distribution might qualify to describe congenital frailty on live birth; we leave this issue open in the present study.

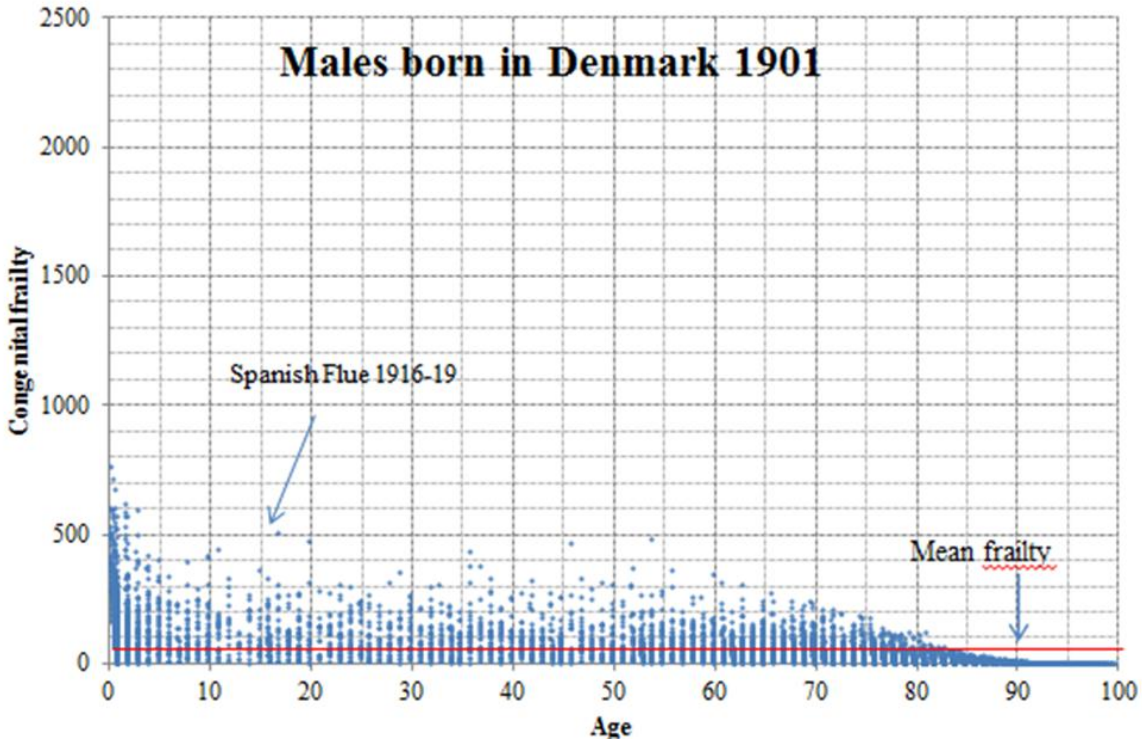
2.3.3 Some predictions under model (1.1)

Selection by congenital frailty and external shocks

Heterogeneity among survivors naturally diminishes across age as individuals, primarily those with congenital frailties above mean, get cropped out by selection. By graphing congenital frailty against individual age at death figure 8 offers straightforward support to this implication of model (1.1). Figure 8 refers to mortality of Icelandic males born around 1767 and of Danish males born in 1901. The Icelandic mortality was recovered

Figure 8 Predicted congenital frailty plotted against age at death
(Cohort size: 10,000)





by Hansen (2004). Both cohorts deviate from the historical situation in two respects. First, to reduce randomness on model-based cohort mortality on population level we consider a birth cohort of 10,000 new born lives; the historical Icelandic birth cohort was much smaller while the Danish cohort size was larger. Second, as the overall trend $\varepsilon(t)$ underestimates mortality somewhat among children and younger persons it has been augmented faintly during peak of the crisis 1784-1786 (Iceland) and 1915-18 (Denmark).

Infant and child mortality takes a terrifying high toll of lives in virtual absence of human control of mortality and that this has dramatic purgative effect on health among the survivors indicated by lower individual z-values on death. In Iceland selection probably intensified dramatically by infectious diseases and various severe multifactor miseries 1784-86 as *indirect* corollaries of extensive earthquakes and severe volcanism 1783-84. However, the environmental disturbances *per se* appear to have instigated few, if any direct casualties (Thoarinsson 1969). Some 22 per cent of the total population perished during the crisis. The cropping out of a great many frail lives during the crisis greatly influenced health and led to abnormally low cohort mortality below the elderly ages in the subsequent epoch.

Compared to the survivor experience of the Icelandic 1767 cohort, mortality had come under considerable human control in the Danish 1901 cohort, not least among infants, children, and younger adults. People with relatively high congenital frailties now tend to live on to much older ages; with likely positive correlation to impaired health and augmenting health costs. Cropping out of frail lives during the Spanish Flue is faintly visible and probably underestimated as the overall trend $\varepsilon(t)$ is a little on the low side as far as children and younger adults are concerned. Despite very different living conditions the frailty distributions conditional on survival to very old ages are nearly the same in the two birth cohorts considered.

Stagnating or temporary deceleration of old-age mortality

That individual health depends on personal genetics and ecological interaction over the life course from conception to death constitutes the fundamental condition of *being*. Modeling the aging process has attracted extensive biological and gerontological focus

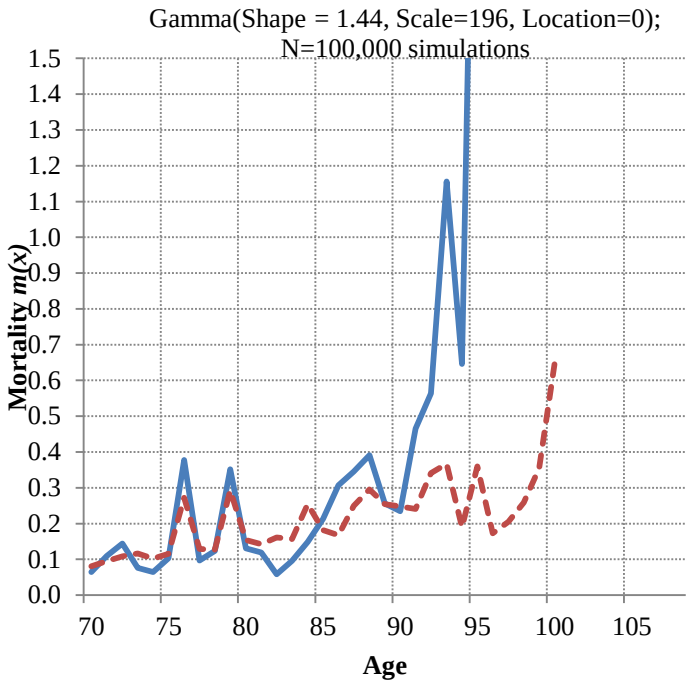
in recent decades. A broad and fairly up-to-date review of this field may be obtained from Handbook of Models for Human Aging edited by Con (2006). A certain transitory stagnation or deceleration of empirical mortality at extreme ages has been known in demography and actuarial mathematics for at least a century. Westergaard (1898) interpreted this phenomenon as a consequence of selection. Referring to recent literature on reliability Leonid and Natalia Gavrilov (2006) draw attention to the similarity in age pattern of the failure rate in living organisms and technical devices. They talk about the closing "period of late life mortality leveling-off" at extreme ages.

Can this extreme age change of mortality be meaningfully and adequately described by parametric statistical approaches? Thatcher et al. (1998) try out their analytic strength using cohort data sampled from thirteen countries. The national cohorts were born between 1871 and 1880. The authors supplement with period data and claim that the chosen thirteen countries have a sufficiently long run of reliable data to "make it possible to assess the relative merits of the various contending models for the way in which the probability of dying changes with age, at least in the range of ages from 80 to 120". How this could ever be accomplished with the given data is not clear as the study does not document empirical population mortality beyond age 105. Furthermore, some of the national birth cohorts appear to be incomplete (left-truncated). The modeling by Thatcher et al. (1998) considers heterogeneity in terms of *age* and *sex* but exclude environmental and all other biological factors but sex and age. The value of including *period* mortality is questionable because of substantial structural mortality change in the course of the demographic transition, cf. for example Appendix A. On this background the conclusion that "no single model was always best" seems pretty meaningless, not least as far as the age segment 105-120 is concerned. The talk ushered by Oeppen and Vaupel (2002) about linear development of period life expectancy with time is likewise empirically unfounded in a long term perspective (Vallin and Mésle 2010). Phantasies about a major future upsurge of life times beyond 105 years are now having their day in the demographic and biological folklore and in political ideologies.

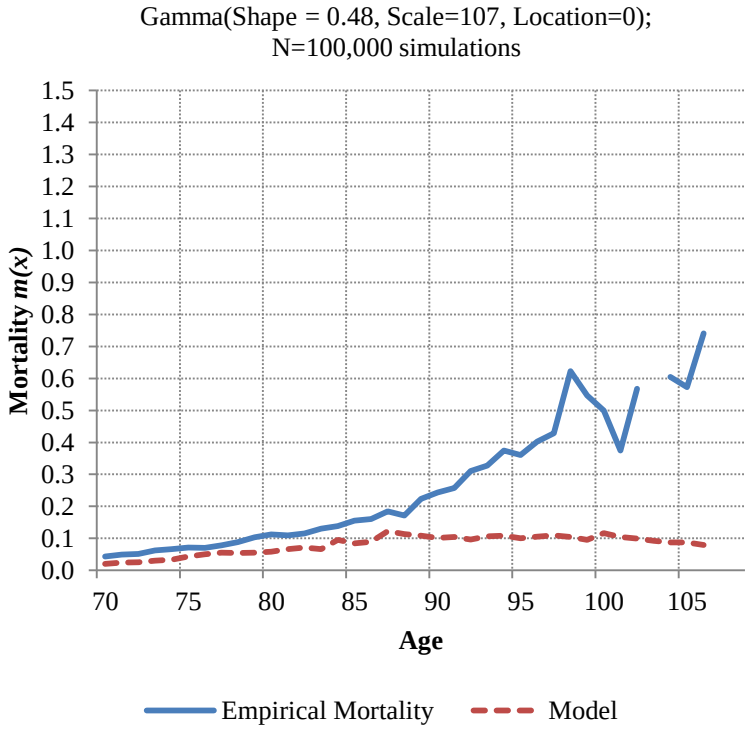
Leonid A. Gavrilov and Natalia S. Gavrilova (2006) discuss shortcomings of parametric statistical modeling to describe aging in general and extreme age mortality change in particular. As an alternative they propose a theoretical statistical systems failure

Figure 9
Empirical and modeled extreme age mortality by age at death
(Cohort size of modeled mortality: 100,000)

Males born in Iceland around 1767



Males born in Denmark 1901



approach; the utility of which remains open as no application is provided in their study or apparently elsewhere.

Featuring selection as proposed by Westergaard (1898) model (1.1) successfully captures basic features of personal genetics and ecological interaction over the life course from conception to extinction of life on firm empirical basis; leaving the intriguing question of an upper limit to human life open. The number of centenarians observed in a birth cohort is a function not only of mortality but perhaps more so of size of the cohort. However, because of very high human mortality in the extreme ages, beyond age 92 say the risk sets of survivors tend to be small and randomness on estimated mortality therefore large; which ideally calls for appropriate statistical approaches on life testing. Obtainable mortality statistics from national statistical agencies are nearly always *approximate* occurrence/exposure rates with rough evaluation of the risk set. Before inclusion into the Berkeley or Human Mortality Database such data appear to have been subject to extensive additional trimming (Wilmoth et al. 2007); which despite all noble intensions makes the quality of extreme age mortality data from this data source questionable. Occurrence-/exposure rates based on micro-simulated personal life times are always founded on exact evaluation of risk time in the present study.

To what does selection over the life course instigate stagnation or deceleration in extreme age mortality? To consider the interpretation advanced by Westergaard (1898) we compare empirical and modeled extreme age mortality of two birth cohorts subject to very different patterns of selection over their life courses viz. Icelandic male cohort born around 1767 and Danish males born in 1901 (figure 8). In both cohorts model (1.1) fit empirical mortality rather closely between age 70 and age 90. Despite increasing problems with empirical data quality as already mentioned some stagnation or deceleration of mortality is evident is evident in both birth cohorts, most pronounced, perhaps, in the Icelandic cohort. Modeled mortality exhibits clear cut stagnation (Iceland, Denmark) and even deceleration (Denmark) beyond age 90. Beyond age 90, modeled mortality is somehow lower than empirical mortality in both examples; possibly because of some undervaluation of baseline mortality.

3 Extension of model (1.1) to describe individual health histories in presence of selection and external influence

Death is the ultimate outcome of somatic disease or violent demise associated with accidents, suicide, or murder. Analysis of mortality by age and cause is commonly based on the competing risks model. We briefly consider some merits and disadvantages of the competing risks model on evaluating change of mortality by cause and global age across the entire life course. Defined on finite state spaces, multivariate stochastic survivor processes may be seen as systems of conditional competing risks models, the term *age* now denoting *biological age* or *seniority in current life state*. The competing risks model opens new vistas for stochastic micro simulation of complex human survivorship in the framework of consistent statistic modeling. We close this paper by sketching how the basic life model (1.1) may be extended to include duration dependent morbidity and transition to death by a given cause in presence of selection by congenital frailty and environment.

3.1 The competing risks model

Keeping sex as a background variable, age-cause specific mortality is normally studied in the framework of the competing risk model on assumption that the net risks $m(x,r)$ by cause r of death are statistically independent; if aggregated over cause-of-death the age-cause specific death risks simply add to overall mortality $m^{\bullet}(x)$ i.e.

$m^{\bullet}(x) = \sum_{\forall r} m(x,r)$. Assuming piecewise constant mortality i.e.

$m(x,r) = m(x+\tau,r)$, $\tau \in [0,t[$, τ denoting length of the interval starting in age x sharp, derivation of the state distribution and all expected life lengths and expected losses of life length become particularly simple and straightforward (Hansen 2007).

Does it make sense to consider net risks? Are net risks identifiable? Such questions are almost philosophical in nature. A general answer is not viable in the bio-social sciences. A net risk $m(x,r) = D(x,r)/A(x,r)$ is defined in absence of all other death risks. A crude risk $q(x,r)dx = \left(D(x,r)/A^{\bullet}(x)\right)dx$ is defined in presence of, and thereby influenced by all other death risks; $A(x,r)$ and $A^{\bullet}(x)dx$ denoting risk sets. It may be

difficult and perhaps impossible to establish empirically whether deaths associated with a given cause *r* are due solely to this very cause or whether there would latent accompanying causes at work. Furthermore, the "true" risk set $A(x,r)$ under exposure to dying from cause *r* escapes observation in most situations. Empirical cause-age specific death rates are normally estimates of crude death risk $q(x,r)dx$ rather than of net risk $m(x,r)$. For a critical review of competing risk models cf. Ferkingstadt (2008).

Lack of homogeneity of the risk set is a serious limitation of the competing risks model if used to evaluating change of overall and age-cause specific mortality across the entire life course (global age). The model is "blind" to mortality differentials linked up with socio-economic behavior e.g. labor force participation which, again, may be correlated with individual genetics and frailty on live birth. Furthermore, the competing risks model does not account for individual exposure to shifting environmental influence on health. Sweeping such heterogeneity under the rug using the classical competing risks model as the salient and one-and-only analytic device may seriously impair insight, not only in socio-economic and various epidemiological aspects of mortality but also of population impacts of artificial immunity. Furthermore, on studying such issues the analyst is commonly faced with bureaucratic walls of access to informative empirical benchmarks, mostly in terms of mediocre life data. Such difficulties call for modeling of better relevance.

3.2 A model of morbidity and mortality by age and cause-of-death

Extending model (1.1) to accommodate health and cause-of-death entails introduction of state space S_1 = {1: Not ill; 2: Ill, diagnosis #h; 3: Dead, diagnosis #k} (Figure 1). Let $q^{ij}(x, x+\Delta_x)$ denote the probability of being in life state j at age $x+\Delta_x$ of someone present in life state i at age x. The relationship between probability $q^{ij}(x, x+\Delta_x)$ and hazard (net risk; force of transition) $m^{ij}(x)$ is then,

$$m^{ij}(x) = \lim_{\Delta_x \to 0} q^{ij}(x, x+\Delta_x)/\Delta_x ; i \neq j; i, j \in \{1,2,3\}; 3i = 3j = 0$$

Let $m^{ij}(x, z_\upsilon)$ indicate the force of transition of individual $\#\upsilon$ with congenital frailty z_υ from life states *i* to *j* in the course of age $[x, x+\Delta_x[$. An extended hazard model defined on state space S_1 suited for stochastic micro simulation may then be stated as follows.

Let person υ be in life state *i* at age *x* sharp. Consider age interval $[x, x+t[$ and let τ denote waiting time to next exit from life state *i* to life state *j* before age $x+t$ with expectation $E(\tau) = \int_0^t P(\tau | \tau < t) d\tau$. Age on exit from life state *i* is then $x + E(\tau)$.

Let the probability of accessing life state *j* before age $x+t$, given presence in life state *i* at age $x+\tau$, be $q^{ij}(x+\tau, x+t)$ where,

$$q^{ij}(x+\tau, x+t) = \int_\tau^t \frac{m^{ij}(x+u)}{m^{i\bullet}(x+u)} q^{i\bullet}(x+u)\, du$$

If for example, probability $q^{ij}(x+\tau, x+t)$ is associated with a given action diagnosis *h*, then probability $q^{ij}(x+\tau, x+t; h)$ may suitably be approached by

$$q^{ij}(x+\tau, x+t; h) = p(h)\, q^{ij}(x+\tau, x+t);$$

statistic $p(h)$ referring to some appropriate empirical probability distribution $H(h)$ of diagnoses.

Let the state of death be labeled by 3. The force of exit from life state *j* from any other cause but 3 is then $m^{j\bullet}(x) - m^{j3}(x, z_\upsilon)$ which could be of any complexity; risk $m^{j3}(x, z_\upsilon)$ as usual referring to the basic frailty model. If cause k of death is independent of a given action diagnosis *h*, death risk $m^{j\delta}(x, z_\upsilon, k)$ may pragmatically be determined as

$$m^{j3}(x, z_\upsilon, k) = p(k)\, m^{j3}(x, z_\upsilon);$$

statistic $p(k)$ referring to some appropriate empirical probability distribution $G(h)$ of causes-of-death. Otherwise death risk $m^{j3}(x, z_\upsilon, k)$ could be stated as contingent on a given action diagnosis *h* i.e. $m^{j3}(x, z_\upsilon, k | h)$.

To accommodate selection in probabilities $q^{ij}(x, x+t); i, j = 1,2, i \neq j,$ consider the following transformations with state labels *1="not ill"* and *2="ill"*,

$$m^{12}(x,t) \exp(z_\upsilon - E[Z])$$

$$m^{21}(x,t)\exp\left(E[Z]-z_{\upsilon}\right)$$

If individual congenital frailty z_{υ} is greater than expected congenital frailty $E[Z]$ then the force of entry into illness will be greater than the mean force of entry; in this case recovery from illness will be delayed by factor $\exp\left(E[Z]-z_{\upsilon}\right)$. If, on the other hand, individual congenital frailty z_{υ} is smaller than expected congenital frailty $E[Z]$ then the force of entry into illness will be smaller than the mean force $z_{\upsilon}=E[Z]$ of entry; in which case recovery from illness will be delayed by factor $\exp\left(E[Z]-z_{\upsilon}\right)$; so persons with high frailties come to prevail in health state "ill" while people with low frailties will tend to stay healthy. Because of selection by congenital frailty, mortality will be higher among ill persons than among healthy persons.

In addition to selection, duration dependency in a transient life state may be influenced by behavioral factors, for example smoking or medical treatment. The feasibility of separating selection from other factors impacting on duration dependency in a transient life state depends heavily on clinical knowledge and the data and empirical benchmarks available. Incorporating such features in the modeling is beyond the scope of this study.

4. Closing remarks

Up to now model (1.1) has been fitted to empirical survivorship of elected birth cohorts from Sweden, Denmark, Iceland, and Japan. Estimating the extended frailty model (1.1) for these birth cohorts presents a number of thorny data problems related to availability; observational plan; level of aggregation in the strategic variables *sex*, *age* and *cause-of-death*; quality and comparability of diagnoses; and impact of technological and environmental change on what people die from.

Various data imperfections related to such issues leave room for much guessing on the demographic and epidemiological center pieces of analytic interest namely the forces of transition in state space S_1 (Figure 1). To make such conjecture informed and enlightening it may be useful to generate representative samples of individual health histories or occurrence/exposure data by stochastic micro-simulation of survivorship in the framework of state space S_1.

References

Bourgeois-Pichat, J. 1946. De la mesure de la mortalité infantile. Population (French edition). 1, 1. 53-68

Bourgeois-Pichat, J. 1951. La mesure de la mortalité infantile. Population (French edition). 6, 2. 233-248

Bourgeois-Pichat, J. 1952. La mesure de la mortalité infantile: II, Les causes de deces. Population (French edition). 6, 3. 459-480

Cox, D.R. (1972). Regression models with life tables (with discussion). Journal of the Royal Statistical Society, B 51, 187-220.

Gompertz, B., (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. Philosophical Transactions of the Royal Society of London, Vol. 115 (1825), 513-585.

Ferkingstadt, Egil (2008). A critical review of competing risk models. Department of Biostatistics, University of Oslo, and Centre for Integrative Genetics, Norwegian University of Life Sciences, 12.2.2008A (Link: sfi.nr.no/sfi/images/e/e4/Trial_ferkingstad.pdf).

Gavrilov, L. A, N. S. Gavrilova(2006). Models of Systems Failure in Aging. In P. Michael Conn (ed., 2006), Handbook of models for human aging, 45-67

Hansen, Hans O. (1978). Some age structural consequences of mortality variations in pre-transitional Iceland and Sweden. In H. Charbonneau & A. Larose (Eds.), The great mortalities: methodological studies of demographic crises in the past, 113-132 (Liège: Ordina Editions).

Hansen, Hans O. (1982a,b). Studies in recent Danish and Greenlandic infant mortality on cohort basis. Main report + Supplement. USI Research Reports No. 86-87.

Hansen, Hans O. (1989). Survivorship and family formation in the parishes of Rødovre and Sejrø from the late 17th century up to the 1801 census. Scandinavian Population Studies 9, 307-35.

Hansen, Hans O. (1996). Social and biological issues in infant survivorship among Danish cohorts born between 1982 and 1990. Yearbook of Population Research in Finland 1996, 82-100.

Hansen, Hans O. (2000). An AIDS model with reproduction - with an application based on data from Uganda. Mathematical Population Studies, 2000, Vol. 8(2), 175-203.

Hansen, Hans O. (2007). Demography 1-2. Lecturing notes. Department of Economics, University of Copenhagen.

Hansen, Hans O. (2008). Issues of Selection in Human Survivorship: A Theory of Mortality Change from the Mid-Eighteenth to the Early Twenty First Century. Discussion Paper No. 2008-18. Department of Economics, University of Copenhagen. (Link: ftp://ftp.ibt.ku.dk/usihoh/Docs)

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

Lee, Ronald D. and Lawrence R. Carter. 1992. Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association 87 (419, September).

Makeham, W.M. 1860. On the law of mortality and the construction of annuity tables. J. Inst. Actuaries 8, 301-310.

Thorarinsson, S. (1969). The Lakagígar eruption of 1783. Journal Bulletin of Volcanology. Volume 33, Number 3. Berlin / Heidelberg: Springer Verlag

Vallin, J., F. Meslé(2010). Espérance de vie : peut-on gagner trois mois par an indéfiniment ? Population et Sociétés Numéro 473 Décembre 2010

Thatcher, A.R, J. W. Vaupel, V. Kannisto, (1998). The Force of Mortality at Ages 80 to 120. Monographs on Population Aging, 5. Odense University Press (Link: **http://www.demogr.mpg.de/Papers/Books/Monograph5/start.htm**)

Vaupel J. W., K.G Manton., E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 1979,16(3), 439-454.

Oeppen J., J.W. Vaupel (2002). Broken Limits to Life Expectancy. Science Magazine 10 May 2002 296 (5, 570): 1029-1031

Westergaard, Harald L. (1898). Oldingealderens Dødelighed (Old age mortality). Natinaløkonomisk Tidsskrift 1898, 349-58.

Wilmoth J.R., K. Andreev, D. Jdanov, and D.A. Glei with the assistance of C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, P. Vachon (2007). **Methods Protocol for the Human Mortality Database** (Link: **http://www.mortality.org/**)

Appendix

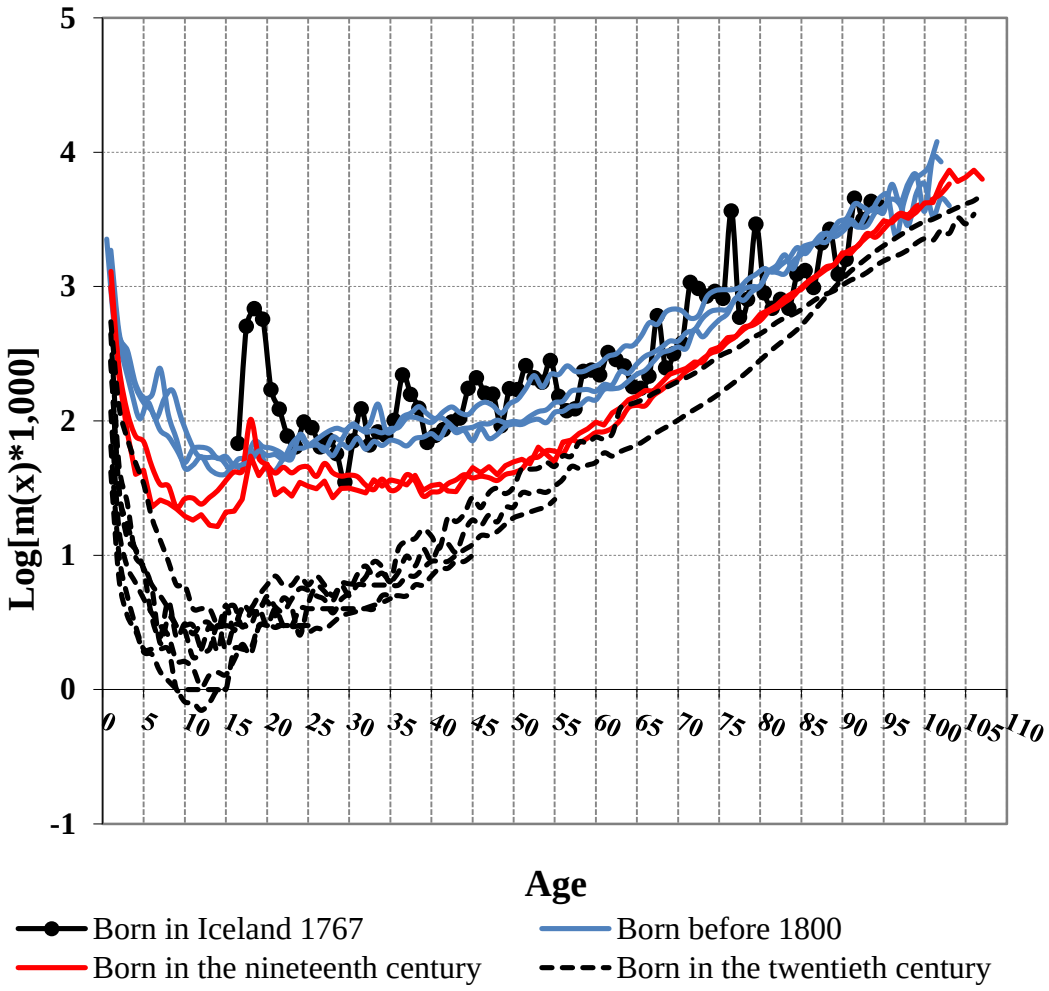
Figure A.1 Empirical mortality of elected female cohorts born before 1802 and in the course of the nineteenth and twentieth centuries (semi-logarithmic scale).

Table A.1 Gamma parameter values of the frailty distributions on live birth while fitting Eq. (1.1) to empirical mortality of the elected birth cohorts.

Table A.2 Demographic results obtained on fitting Eq. (1.1) to the empirical mortality of the elected birth cohorts.

For more results cf. ftp://ftp.ibt.ku.dk/usihoh/Selection in human survivorship/, including the PowerPoint presentation of this study, and Hansen (2008).

Figure A.1. Empirical mortality of elected female cohorts born before 1800 and in the course of the nineteenth and twentieth centuries (semi-logarithmic scale)



Note

- Cohorts born before 1802:
  Iceland 1767;
  Sweden 1751, 1801
- Cohorts born between 1802 and 1900:
  Denmark 1835 and 1851;
  Sweden 1851
- Cohorts born in the twentieth century:
  Sweden and Denmark 1901, 1944, 1952;
  Japan 1950, 1960, 1970, 1980, and 1990

Sources

Sweden, and Denmark

- Berkeley Mortality Data Base

Iceland

- Icelandic mortality recovered by Hansen 2004.

Japan

- Berkeley Mortality Data
- Base Statistics and Information Department, Minister's Secretariat, Ministry of Health, Labor and Welfare "Vital Statistics“ (Japan)
- Statistics Bureau, the Director-General for Policy Planning (Statistical Standards) and the Statistical Research and Training Institute (Japan)

Table A1

Empirical and model-based mortality estimates of elected birth cohorts[1]

Birth cohort	Age span $[x,y[$	Empirical estimates			Model-based estimates		
		Mean frailty $\mathbb{1}$	Infant mortality $m[0,1[$	Life expectancy $e[x,y[$	Mean frailty $\mathbb{1}$	Infant mortality $m[0,1[$	Life expectancy $e[x,y[$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Sweden, males							
1751	0-100	90.32	0.238	33.8	125.6	0.244	34.5
1801	0-100	97.74	0.258	36.2	128.8	0.254	35.8
1851	0-100	70.84	0.187	43.9	101.2	0.195	45.3
1901	0-100	45.88	0.121	56.8	57.4	0.110	57.1
1944	0-55	14.27	0.038	52.8	14.2	0.029	51.8
1952	0-47	9.29	0.022	45.0	14.3	0.028	45.0
Sweden, females							
1751	0-100	90.32	0.238	39.8	97.0	0.191	40.2
1801	0-100	97.74	0.258	40.6	119.3	0.231	40.3
1851	0-100	70.84	0.187	47.5	82.1	0.159	47.4
1901	0-100	45.88	0.121	61.8	48.2	0.096	59.7
1944	0-55	14.27	0.038	52.2	8.2	0.016	53.4
1952	0-47	7.32	0.017	46.7	5.9	0.010	46.7

[1] All model-based estimates relate to the basic frailty model; cf. Eq. (1.1).

Birth cohort	Age span $[x, y[$	Empirical estimates			Model-based estimates		
		Mean frailty $\mathbb{I}$	Infant mortality $m[0,1[$	Life expectancy $e[x, y[$	Mean frailty $\mathbb{I}$	Infant mortality $m[0,1[$	Life expectancy $e[x, y[$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Denmark, males							
1835	0-100	89.81	0.237	42.2	112.4	0.220	40.4
1851	0-100	92.46	0.219	43.2	106.8	0.205	41.9
1901	0-100	62.65	0.165	56.3	51.4	0.101	62.2
1944	0-55	21.68	0.057	50.2	14.0	0.029	51.9
1952	0-47	14.07	0.033	45.4	14.6	0.032	44.6
Denmark, females							
1835	0-100	70.67	0.186	45.3	106.8	0.209	42.2
1851	0-100	78.53	0.186	45.0	100.0	0.193	43.2
1901	0-100	49.18	0.130	61.7	34.7	0.100	65.7
1944	0-55	16.83	0.044	52.3	10.2	0.020	53.0
1952	0-47	10.76	0.026	46.2	8.82	0.015	46.0
Iceland, males							
1767	16-95	*	*	32.7	282.2	0.496	38.0
	0-95	*	*	*	282.2	0.496	26.6
Modified trend							
	16-95	*	*	*	282.2	0.496	34.7
	0-95	*	*	*	282.2	0.496	24.6

Birth cohort	Age span $[x, y[$	Empirical estimates			Model-based estimates		
		Mean frailty $\mathbb{1}$	Infant mortality $m[0,1[$	Life expectancy $e[x, y[$	Mean frailty $\mathbb{1}$	Infant mortality $m[0,1[$	Life expectancy $e[x, y[$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Iceland, females							
1767	16-95	*	*	38.3	246.8	0.440	39.3
	0-95	*	*	*	246.8	0.440	28.3
Modified trend							
	16-95	*	*	*	246.8	0.440	36.0
	0-95	*	*	*	246.8	0.440	26.8
Japan, males							
1950	0-54	23.59	0.062	49.9	23.6	0.062	49.8
1960	0-44	12.97	0.034	42.6	13.0	0.033	42.5
1970	0-34	5.80	0.015	34.1	5.8	0.015	34.1
1980	0-24	3.11	0.008	24.7	3.1	0.007	24.7
1990	0-14	1.89	0.005	14.9	1.9	0.005	14.9
Japan, females							
1950	0-54	20.74	0.055	49.9	20.7	0.053	50.0
1960	0-44	10.62	0.028	43.2	10.6	0.027	43.0
1970	0-34	4.45	0.012	34.4	4.4	0.012	34.3
1980	0-24	2.47	0.007	24.8	2.5	0.006	24.8
1990	0-14	1.58	0.004	14.9	1.6	0.004	14.9

Note. Symbol * denotes statistic undefined (missing data)

Table A2

Parameter values and frailty related to model-based heterogeneity of mortality of Elected cohorts[1]

Birth cohort	Gamma statistics				Sum of squared deviation	
	Shape α	Scale β	Est. $E[Z]$ $\alpha\beta$	Est. $VAR[Z]$ $\alpha\beta^2$	Age span [x,y[	SSD[x,y[
1	**2**	**3**	**4**	**5**	**6**	**7**
Sweden, males						
1751	1.57	80	125.6	10048	0-90	0.0967
1801	1.48	87	128.8	11202	0-90	0.0434
1851	0.88	115	101.2	11638	0-90	0.0177
1901	0.70	82	57.4	4707	0-90	0.0120
1944	1.09	13	14.2	184	0-50	0.0002
1952	0.34	42	14.3	600	0-50	0.0001
Sweden, females						
1751	1.47	66	97.0	6403	0-90	0.0219
1801	0.97	123	119.3	14675	0-90	0.0167
1851	1.14	72	82.1	5910	0-90	0.0175
1901	0.73	66	48.2	3180	0-90	0.0156
1944	1.02	8	8.2	65	0-50	0.0002
1952	0.42	14	5.9	82	0-50	0.0001

Birth cohort	Gamma statistics				Sum of squared deviation	
	Shape α	Scale β	Est. $E[Z]$ $\alpha\beta$	Est. $VAR[Z]$ $\alpha\beta^2$	Age span [x,y[	SSD[x,y[
1	**2**	**3**	**4**	**5**	**6**	**7**
Denmark, males						
1835	1.07	105	112.4	11797	0-90	0.0243
1851	0.98	109	106.8	11643	0-90	0.0250
1901	0.48	107	51.4	5496	0-90	0.0538
1944	1.00	14	14.0	196	0-50	0.0012
1952	1.22	12	14.6	176	0-50	
Denmark, females						
1835	0.98	109	106.8	11643	0-90	0.0194
1851	1.00	100	100.0	10000	0-90	0.0219
1901	0.62	56	34.7	1944	0-90	0.0078
1944	1.28	8	10.2	82	0-50	0.0004
1952	1.26	7	8.82	62	0-50	
Iceland, males						
1767	1.44	196	282.2	55319	20-90	0.1779
					0-95	*
Modified trend						
	1.44	196	282.2	55319	20-90	0.1411
					0-95	*

Birth cohort	Gamma statistics				Sum of squared deviation	
	Shape α	Scale β	Est. $E[Z]$ $\alpha\beta$	Est. $VAR[Z]$ $\alpha\beta^2$	Age span [x,y[	SSD[x,y[
1	**2**	**3**	**4**	**5**	**6**	**7**
Iceland, females						
1767	1.41	175	246.8	43181	20-90	0.3110
			246.8		0-95	*
Modified trend						
			246.8		20-90	0.1383
			246.8		0-95	*
Japan, males						
1950	1.20	20	23.6	464	0-49	0.000083
1960	0.55	24	13.0	306	0-39	0.000099
1970	1.15	5	5.8	29	0-29	0.000040
1980	1.15	3	3.1	8	0-19	0.000029
1990	0.90	2	1.9	4	0-10	0.000020
Japan, females						
1950	1.15	18	20.7	374	0-49	0.000077
1960	0.55	19	10.6	205	0-39	0.000064
1970	0.90	5	4.4	22	0-29	0.000027
1980	0.45	5	2.5	14	0-19	0.000017
1990	0.45	4	1.6	6	0-10	0.000007

Note. Symbol * denotes statistic undefined (missing data)